Coverage Path Planning for AUVs Cooperative Environment Detection in Integrated Underwater Acoustic Communication and Detection Networks

Yike Wu[⋆‡], Xiaoxiao Zhuo[⋆], Liang Tang[⋆], Fengzhong Qu[∗], Yu Zhao[⋆], and Zhiyong Bu[⋆]
Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China[⋆]
University of Chinese Academy of Sciences, Beijing 101408, China[‡]
Zhejiang Key Laboratory of Ocean Observation-Imaging Testbed, Zhejiang University, Zhoushan 316000, China[∗]
Email: {wuyk, zhuoxx, liang.tang, yu.zhao, zhiyong.bu}@mail.sim.ac.cn[⋆], jimqufz@zju.edu.cn[∗]

***Abstract*—In this paper, we investigate the coverage path planning (CPP) scheme for autonomous underwater vehicles (AUVs) cooperative environment detection in integrated underwater acoustic communication and detection networks (UCDNs), where multiple AUVs detect unexplored oceanic environments and avoid obstacles. Firstly, we present the detection range prediction model related to oceanic environmental parameters and propose the detection and communication scheme in UCDNs. Secondly, to conduct the cooperative environment detection mission, we formulate the CPP problem as a mixed combinatorial and sequential quadratic optimization problem to maximize the coverage ratio and minimize the path length of AUVs. To solve this problem, we investigate the multi-agent proximal policy optimization (MAPPO)-based CPP scheme. In specific, the CPP problem is modeled as a partially observable Markov decision process (POMDP). Since the path planning of the AUVs is not only related to the local information but also the other AUVs' information, the information should be shared among AUVs based on the UCDNs. Furthermore, we introduce the MAPPO-based algorithm under the centralized training with decentralized execution (CTDE) architecture. Extensive simulations are carried out to demonstrate the strength of the proposed scheme.**

I. Introduction

With the great prosperity of marine science and technology, autonomous underwater vehicles (AUVs) have become more and more sophisticated, and multiple AUVs are expected to collaboratively accomplish complex and larger-scale missions. AUVs cooperative complete coverage path planning (CPP) mission has wide application prospects, such as subsea exploration, underwater search, and submarine survey [1], [2]. In general, the CPP task requires multiple AUVs to detect the interest of area while avoiding collision with unforeseen obstacles. Therefore, communication and detection are two indispensable functions to enable information sharing among AUVs and environment detection. Recently, integrated underwater acoustic detection and communication networks (UCDNs) have been developed and envisioned as a promising network for providing both communication and detection services for underwater applications [3], [4]. Empowered with UCDNs, AUVs can implement CPP tasks flexibly and lightweightly.

✉Liang Tang (liang.tang@mail.sim.ac.cn) is the corresponding author of this paper.

However, AUVs cooperative CPP tasks are challenged by limited communication resources, complex marine physics, and unforeseen obstacles. Firstly, the obtained detection information should be shared among multiple AUVs promptly. Secondly, to complete the area coverage detection, the detection ability of AUVs is affected by marine physics due to the characteristic of underwater acoustic propagation. Thirdly, the path planning of the AUV is expected to avoid unforeseen obstacles online and adaptively.

Several existing works are proposed to design the CPP scheme for AUVs, which can be divided into two aspects, static global offline algorithms and dynamic local online algorithms. Regarding offline algorithms, they require accurate modeling of the entire underwater environment before coverage path planning is initiated [5]. On the other hand, online algorithms can independently determine path planning based on limited observed environmental information. Efforts to enhance the efficiency of multi-AUV cooperative CPP schemes mainly fall into two categories: centralized and decentralized cooperative CPP methods. For centralized cooperative CPP algorithms, a center AUV will carry out the task allocation scheme and divide the large task region into several subtasks for coverage [6]. However, due to environmental uncertainties, this approach may lead to unfair distribution and encounter limitations in communication range. To address fairness concerns, a K-means based dynamic cooperative partition strategy is introduced to minimize detection range overlap and ensure equal workload [7]. Furthermore, the task is allocated based on the dot-spreading-based mission assignment scheme and AUV plans path based on the virtual attraction-based CPP in [8]. In terms of the decentralized coverage planning scheme, each AUV independently makes decisions based on limited environmental information. To achieve fully decentralized collaborative exploration, a decentralized reinforcement learning framework with dual guidance (DODGE) is proposed to integrate multiple agents' information into the observation environment [9]. In existing works, two aspects are not simultaneously considered, i.e., the impact of dynamically changing ocean environments on AUV detection performance and the robustness of algorithms adapting to variable underwater en-

979-8-3503-8741-4/24/$31.00 ©2024 IEEE

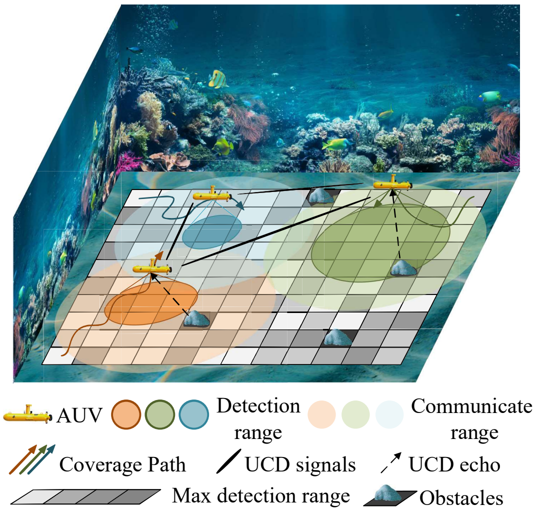


Fig. 1: System model.

vironments.

In this paper, we propose a CPP scheme for AUVs cooperative environment detection in UCDNs. We first introduce the communication and detection model to coordinate the detection and information sharing among AUVs in UCDNs. Specifically, a physical environment-based underwater detection model is proposed to predict the detection range. Secondly, we formulate the CPP problem as a mixed combinatorial and sequential quadratic optimization problem to maximize the coverage ratio and minimize the path length of AUVs. Furthermore, to plan the adaptive and flexible path, we model the CPP problem as the partially observable Markov decision process (POMDP) and investigate the multi-agent proximal policy optimization (MAPPO)-based CPP algorithm to solve the problem. Simulations are carried out to evaluate the performance of the proposed algorithm.

The contributions of this paper are summarized as follows:

1) The underwater acoustic communication and detection model in dynamically changing oceanic environments for UCDNs is presented.
2) We formulate the CPP problem and investigate the MAPPO-based CPP algorithm, which enables the AUV to achieve the large coverage ratio and short path length adaptive to various oceanic environments.

The remainder of this paper is organized as follows. Section II presents the system model. In Section III, we formulate the problem. Following this, the proposed scheme is illustrated in Section IV. Section V evaluates the performance of the proposed scheme. Finally, Section VI summarizes the conclusions of this paper.

II. System Model

A. Network Model

In this paper, we consider the scenario in which multiple AUVs collaborate to plan coverage paths to explore unknown underwater environments as well as detect the surrounding environment to avoid any obstacle collision, as illustrated in Fig. 1. The network consists of the coverage target area and K AUVs, denoted by $\mathcal{K} = \{A_1, A_2, \cdots, A_K\}$.

- Coverage Target Area: The coverage target area is a two-dimensional marine space divided into an $L \times W$ grid map. Before detection, the physical oceanic environment parameters sampled by the ocean scientific instruments have been obtained. Based on the detection range prediction model illustrated afterward in Section II. B, all AUVs initially possess the detection ability information $\mathcal{M}$ at each position. During the travel, AUVs at the position $\mathbf{p}_k^t = (l_k^t, w_k^t)$ at time t will explore the area $\mathcal{P}_k^{d,t}$ within the maximum detection range $r_k^{d,t}$, i.e., $\mathcal{P}_k^{d,t} = \{\mathbf{p}_{k,1}^{d,t}, \mathbf{p}_{k,2}^{d,t}, \cdots, \mathbf{p}_{k,N}^{d,t}\}$ when $||\mathbf{p}_{k,n}^{d,t} - \mathbf{p}_k^t||_2 \leq r_k^{d,t}$. Here, $\mathbf{p}_{k,n}^{d,t}$ represents position coordinate $(l_{k,n}^{d,t}, w_{k,n}^{d,t})$. As such, the detection ability information of the area $\mathcal{M}$ for AUV A_k will be updated by $\mathcal{M}(\mathbf{p}_k^t) = \mathcal{M}(l_k^t, w_k^t) = \mathcal{P}_k^{d,t}$.
- AUVs: All AUVs are equipped with the same integrated underwater acoustic communication and detection devices. Nevertheless, since the underwater acoustic propagation characteristic is highly related to the environment, the communication and detection capacities of AUVs will change with the ocean environment parameters, such as temperature, salinity, and depth. In UCDNs, AUVs send the integrated underwater acoustic communication and detection signal (UCD) to detect the environment and share information among AUVs. Then AUVs receive the echo signal to detect the environment. Specifically, for the CPP tasks in UCDNs, the detection scheme and communication scheme will be presented as follows.

B. Detection Scheme in UCDNs

1) Detection Range Prediction Model: For the active detection using UCD signals, the detection range is related to the transmission loss (TL) of the sound waves denoted by $TL(\mathbf{r}, env)$, which can be calculated by the sound tool, for example, range-dependent acoustic modeling (RAM) [10]. The env is determined by the sound-speed profile (SSP) and other environmental parameters. To calculate the max detection range r_k^d in a position of AUV A_k, the threshold of the transmission loss should be calculated, which is defined as the figure of merit (*FoM*), i.e., $TL(r_k^d, env) = FoM_A$. Furthermore, the *FoM* can be calculated by the active sonar equation as follows [11]

$$FoM_A = \frac{1}{2}(SL + TS - (NL_S - DI) - DT_S), \quad (1)$$

where *SL* is the source level of the transmitted UCD signal, *TS* is the target strength, NL_S is the ambient noise level related to the noise of turbulence, shipping, and waves, *DI* is the directivity index, and DT_S is the detection signal-to-noise ratio (SNR) threshold of received echo signal [12].

2) Target Detection Scheme: To detect the environment, the joint mono-static and multi-static active detection scheme is applied. As illustrated in Fig. 2, each AUVs send the UCD signal periodically and receive the echo signal. Based on the echo signal, not only the unknown underwater environments

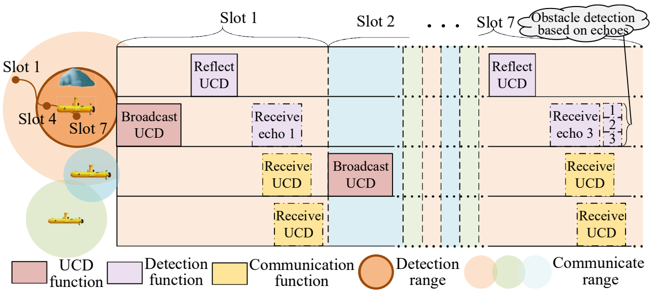


Fig. 2: Timeline of communication and detection mechanism.

can be explored but also the existence and locations of obstacles within the detection range can be obtained.

C. Communication Scheme in UCDNs

Multiple AUVs cooperatively complete the coverage task and make the decision on their path planning. To prevent them from redundantly covering areas already covered by other AUVs, the effective information-sharing scheme should be maintained. As shown in Fig. 2, the AUVs send the UCD signal containing the information about their locations and their action a_k^t, other AUVs will receive the UCD signal and obtain the information. When AUV A_k receives the UCD signal broadcast by cooperating AUV $A_{k'}$ containing $a_{k'}^t$, it will update the detected map information based on the detection ability information M_k^E.

III. Problem Formulation

To complete the area coverage mission, the objective of the proposed cooperative CPP scheme is to maximize the coverage efficiency R_t, which is defined as the coverage ratio E_t divided by the coverage path length d_t, i.e., $R_t = E_t/d_t$.

At each time unit, AUVs move a grid length Δd. Then the coverage path length of all AUVs in period time T_k^t can be calculated as follows:

$$d_t = \sum_{t=1}^{T_k^t-1} \sum_{k=1}^{K} \Delta d \sqrt{(x_k^{t+1} - x_k^t)^2 + (y_k^{t+1} - y_k^t)^2}. \tag{2}$$

The coverage ratio is defined as the percentage of the explored target area accounting for the whole target area, which is the sum of the coverage ratio of change Δe_k^t in each time unit during the AUV A_k trajectory period time T_k^t, which can be expressed as

$$E_t = \sum_{t=1}^{T_k^t-1} \sum_{k=1}^{K} \Delta e_k^t. \tag{3}$$

Here, Δe_k^t is related to the detection range of positions along the AUVs' path and can be expressed by

$$\Delta e_k^t = \frac{\|\mathbf{O}_k^t\|_0 - \|\mathbf{O}_k^{t-1}\|_0}{L \times W}, \tag{4}$$

Here, $\mathbf{O}_k^t$ is a two-dimensional $L \times W$ matrix representing the observation of the target area for AUV A_k at time t, in which the element $O_k^t(l, w)$ is denoted by

$$O_k^t(l,w) = \begin{cases} 0.3\ , & \text{detected when } (l,w) \in \bigcup_{t'=1}^{t} \mathcal{P}_k^{d,t'}, \\ 0.6\ , & \text{position of AUVs when } (l,w) = p_k^t, \\ 1\ , & \text{obstacles}, \\ 0\ , & \text{undetected, otherwise}. \end{cases} \tag{5}$$

When AUVs move along the travel trajectory at the position $\mathbf{p}_k^t = (x_k^t, y_k^t)$, the area $\mathcal{P}_k^{d,t}$ within the detection range will be detected. Then the observation of the target area $O_k^t(l,w)$, $\forall \mathbf{p}_{k,n}^{d,t} = (l,w) \in \mathcal{P}_k^{d,t}$ will be updated. As the AUV moves, a new batch of grids will be detected, and the area of the undetected region in the status matrix $\mathbf{O}_k^t$ becomes smaller, eventually completely covering the target area.

To complete the CPP tasks, the following constraints must be satisfied.

1) All the target areas should be detected when the CPP tasks end within period time T, i.e.,
$$E = \sum_{t=1}^{T-1} \sum_{k=1}^{K} \Delta e_k^t = 1. \tag{6}$$
2) The trajectories of AUVs should avoid all obstacles, i.e.,
$$O^k(p_k^t) \neq 1,\ \forall p_k^t = (x_k^t, y_k^t) \in (\mathcal{X}_k, \mathcal{Y}_k), t \in T, k \in K, \tag{7}$$
where $(\mathcal{X}_k, \mathcal{Y}_k)$ represent the trajectories of AUV A_k.
3) All AUVs need to travel within the target area, i.e.,
$$1 \le x_k^t \le W, 1 \le y_k^t \le L, \forall (x_k^t, y_k^t) \in (\mathcal{X}_k, \mathcal{Y}_k). \tag{8}$$
4) AUV can only move to four directions at each step, i.e.,
$$|x_k^t - x_k^{t-1}| + |y_k^t - y_k^{t-1}| = 1, 0 < t \le T, \forall k \in K. \tag{9}$$
5) AUVs start and stop the CPP task simultaneously, i.e.,
$$T_k = T,\ \forall k \in K. \tag{10}$$

To achieve the most efficient coverage process and ultimately achieve full coverage of the target area, the optimization problem can be formulated as follows:

$$\max_{x_k^t \in \mathcal{X}_k, y_k^t \in \mathcal{Y}_k} \quad \frac{\sum_{t=1}^{T_k^t-1} \sum_{k=1}^{K} \Delta e_k^t}{d_t} \tag{11a}$$
$$\text{s.t.} \quad \sum_{t=1}^{T-1} \sum_{k=1}^{K} \Delta e_k^t = 1, \tag{11b}$$
$$O^k(p_k^t) \neq 1,\ \forall p_k^t \in (\mathcal{X}_k, \mathcal{Y}_k), t \in T, k \in K, \tag{11c}$$
$$1 \le x_k^t \le W, 1 \le y_k^t \le L, \forall (x_k^t, y_k^t) \in (\mathcal{X}_k, \mathcal{Y}_k), \tag{11d}$$
$$|x_k^t - x_k^{t-1}| + |y_k^t - y_k^{t-1}| = 1, \forall t \in T, k \in K, \tag{11e}$$
$$T_k = T,\ \forall k \in K. \tag{11f}$$

In the problem, the objective (11a) maximizes the coverage efficiency of each step during the path planning of all AUVs. This problem can be regarded as a mixed combinatorial and sequential quadratic optimization problem. To solve this problem adaptively and online, we investigate the CPP problem through multiagent reinforcement learning (MARL) algorithm.

IV. Proposed MAPPO-based CPP Algorithm

In the multi-agent system, each AUV is an independent agent that observes the local state and makes the path planning decision based on its own policy to maximize a local reward. Therefore, we formulate the CPP problem as a multi-agent partially observable Markov decision process (POMDP), which consists of state, action, observation, and reward, i.e.,

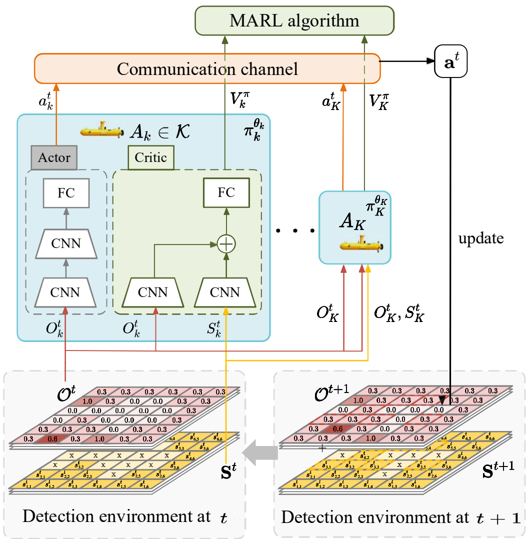


Fig. 3: MAPPO training architecture.

$\left(\{\mathcal{S}_k\}_{k\in\mathcal{K}}, \{\mathbf{a}_k\}_{k\in\mathcal{K}}, \{\mathcal{O}_k\}_{k\in\mathcal{K}}, \{J_k\}_{k\in\mathcal{K}}\right)$. We will illustrate the definitions of each element in MARL for an agent.

State: The state set $\mathcal{S}_k^t$ is a two-dimensional $L \times W$ matrix representing the partially observable detection status of the target area for AUV A_k at position p_k^t, in which the element $s_k^t(p_k^t)$ is denoted by

$$s_k^t(p_k^t) = \begin{cases} 0\ , & \text{undetected}, \\ 1\ , & \text{detected}. \end{cases} \tag{12}$$

Action: The action set of AUV A_k is $\mathbf{a}_k = \{a_k^1, a_k^2, \cdots, a_k^T\}$, where the action a_k^t is the path planning action at time t to visit the grid with four actions, i.e., left, right, top, and bottom.

Observation: The matrix $\mathbf{O}_k^t$ is the observations for the partially observable state S_k^t according to the definition of O_k^t in (5) in Section III.

Reward: The reward of each agent should be designed to achieve the objective of the formulated CPP problem. Firstly, to complete the CPP problem, the large constant reward j_0 is set for an agent to complete the area coverage (constraint (11b)). Secondly, to avoid obstacles and travel within the target area (constraints (11c)and (11d)), the penalty p_0 is set. Thirdly, to complete the task with a large coverage ratio and short path length (objective (11a)), the coverage ratio is rewarded and the path length is punished with coefficient j_1 and p_1, respectively. Therefore, the reward $J(S_k^t, a_k^t) \in \mathcal{J}_k$ is associated with the transition from S_k^t to S_k^{t+1} under action a_k^t with observation O_k^t, i.e.,

$$J(S_k^t, a_k^t, O_k^t) = \begin{cases} j_0, & \sum_{t=1}^{T-1}\sum_{k=1}^{K} \Delta e_k^t = 1, \\ p_0, & O_k^t(l, w) = 1, \\ j_1 \Delta e_k^t - p_1 \Delta d, & \text{otherwise}. \end{cases} \tag{13}$$

Since the MDP of an AUV is influenced not only by the policy of detection ability in the map but also by the maps explored by its neighboring AUVs, the training processes of multiple agents are not entirely independent. We propose a MARL algorithm named MAPPO to optimize the CPP scheme by extending PPO based on centralized training with decentralized execution (CTDE) in the multiagent environment, in

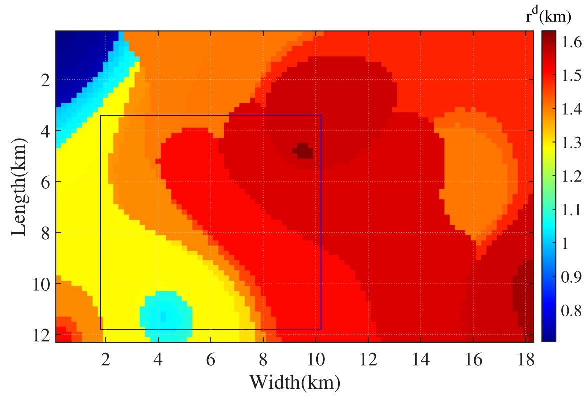


Fig. 4: The detection range.

which each AUV determines the action based on local state, observation, and rewards but shares the action with each other cooperatively to train the centralized critic network [13].

The architecture of the MAPPO-based CPP scheme is shown in the Fig. 3. In the local actor-network, at time t, based on the partial observation O_k^t, the AUV can obtain the local reward as $J_k^t = J(S_k^t, a_k^t, O_k^t)$. Then the AUV A_k would perform an action a_k^t. Note that the action is shared with other AUVs via communication links. The virtual center AUV will update the observation and state based on the joint action $\mathbf{a}^t$ of all AUVs with the state transition probability function $P(\mathbf{S}^{t+1}|\mathbf{S}^t, \mathbf{a}^t)$ and observation function (5). In the centralized critic network, the policy of each AUV is improved through gradient ascent in MAPPO algorithm based on the joint local state and observation. More specifically, the local reward J_k^t of each AUV A_k and discount factor $\gamma \in [0, 1)$ get the discounted return $G_t^i = \sum_{l=0}^{\infty} \gamma^l r_k^{t+l}$. Then it can used to estimate the value of each AUV A_k as the centralized value functions $V_k^\pi(S_k^t) = \mathbb{E}^t[G_k^t \mid S_k^t]$ and the corresponding action-value functions $Q_k^\pi(S_k^t, \mathbf{a}^t) = \mathbb{E}^t[G_k^t \mid S_k^t, \mathbf{a}^t]$. Then the advantage functions are given by $A_k(S_k^t, \mathbf{a}^t) = Q_k^\pi(S_k^t, \mathbf{a}^t) - V_k^\pi(S_k^t)$. Finally, the policy gradient of AUV A_k can be expressed as

$$g_k = \mathbb{E}^t\left[\nabla_\theta \log \pi_k^{\theta_k}\left(a_k^t \mid O_k^t\right) \hat{A}_k\left(S_k^t\right)\right]. \tag{14}$$

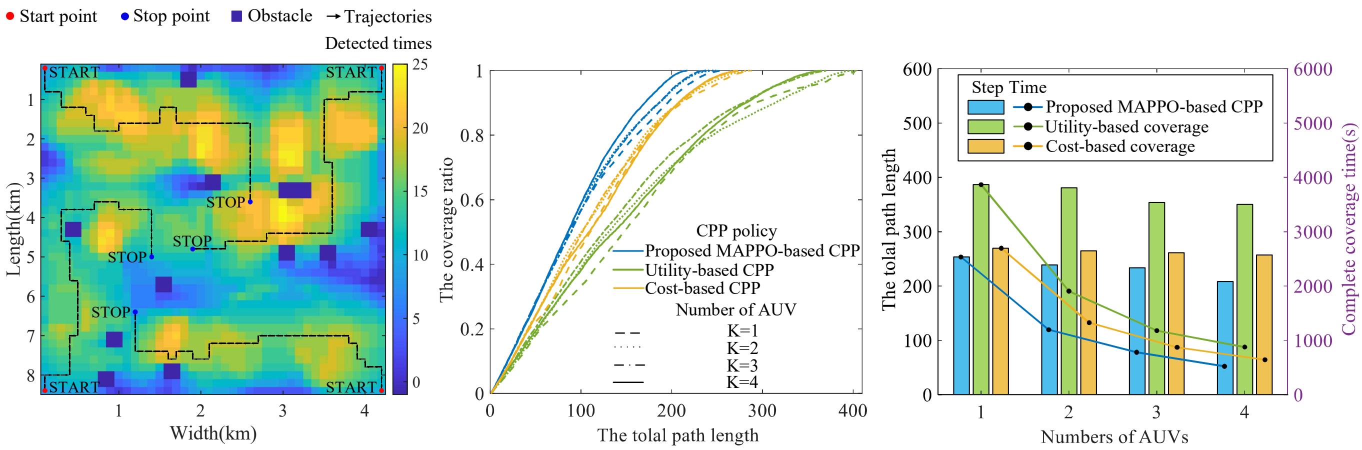
Here, the joint policy gradient then would be used in MAPPO algorithm to improve and update the joint policy $\boldsymbol{\pi}_\theta(\mathbf{a}^t \mid \mathbf{O}^t) = \prod_{k\in\mathcal{K}} \pi_{\theta_k}^t(a_k^t \mid O_k^t)$ with $\boldsymbol{\theta} = \{\theta_k\}_{k\in\mathcal{K}}$.

V. Performance Evaluation

Simulations are carried out to demonstrate the performance of the proposed MAPPO scheme in terms of the total path length, complete coverage time, and coverage ratio with different numbers of AUVs. Two non-learning-based CPP schemes are compared. In the cost-based CPP scheme, AUVs choose an action based on the travelling cost from the nearest frontier cell [14]. In the utility-based CPP scheme, the AUV takes action based on the updated global distance utility function satisfy all the criteria [15].

A. Simulation Setup

In this simulation, to adapt to the ocean parameters, the SSP and other marine parameters obtained from marine experiments are imported into MATLAB Acoustics Toolbox RAM function to simulate the transmission loss. Then the detection



(a) An example of CPP with 4 AUVs.

(b) The coverage ratio with the path length.

(c) The path length and complete coverage time.

Fig. 5: Simulation result

range is calculated and shown in Fig. 4. The coverage task area is obtained from Fig. 4 with range 8.4 km × 8.4 km and divided into grids with range 200 m. All AUVs maintain a speed of 10 m/s. The number of AUVs varies from 1 to 4. In the Monte-Carlo simulations, the distribution of obstacles in the map changes with each cycle of execution.

B. Simulation Results

Fig. 5(a) displays an example of CPP results based on the proposed MAPPO-based CPP scheme. It shows that AUVs can accomplish CPP with obstacle avoidance. The heatmap represents the number of times that the area has been detected.

As depicted in Fig. 5(b), the coverage ratio changes with the total path length of AUVs and the number of AUVs is compared. Results show that with the increase of the total path length of AUVs, the coverage ratio of the proposed MAPPO-based CPP scheme changes larger and faster than that of the other two schemes. In addition, with the increase of the number of AUVs, the coverage ratio will increase stably in the MAPPO-based scheme while unstable in the two other schemes.

In terms of the path length and task time, the influence of the number of AUVs is shown in Fig. 5(c). Results show that with the increase of the number of AUVs, the path length is reduced slightly and the complete coverage time is greatly reduced. In addition, the path length and complete coverage of the proposed MAPPO-based CPP scheme is shorter than the other two methods.

VI. Conclusion

In this paper, we have presented a cooperative CPP scheme to detect unexplored oceanic environments while avoiding obstacles in UCDNs. Furthermore, we have formulated the CPP problem to maximize the coverage ratio and minimize the path length of AUVs and proposed the MAPPO-based CPP scheme to solve it. Extensive simulations have shown the advantage of the performance of the proposed scheme. For the future work, we will study the joint optimization of AUVs' path and the detection performance.

Acknowledgment

This work was supported by XTCX-KJ-2022-2-02.

References

[1] S. W. Cho, H. J. Park, H. Lee, D. H. Shim, and S.-Y. Kim, "Coverage path planning for multiple unmanned aerial vehicles in maritime search and rescue operations," *Comput. Ind. Eng.*, vol. 161, pp. 107 612–107 627, Nov. 2021.

[2] C. Lin, G. Han, T. Zhang, S. B. H. Shah, and Y. Peng, "Smart underwater pollution detection based on graph-based multi-agent reinforcement learning towards AUV-based network ITS," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7494–7505, Jul. 2023.

[3] S. M. Murphy, J. G. E. Scrutton, and P. C. Hines, "Experimental implementation of an echo repeater for continuous active sonar," *IEEE J. Ocean. Eng.*, vol. 42, no. 2, pp. 289–297, Apr. 2017.

[4] Y. Wang, Z. Shi, X. Ma, and L. Liu, "A joint sonar-communication system based on multicarrier waveforms," *IEEE Signal Process. Lett.*, vol. 29, pp. 777–781, Apr. 2022.

[5] J. Yang, J. Ni, M. Xi, J. Wen, and Y. Li, "Intelligent path planning of underwater robot based on reinforcement learning," *IEEE Trans. Autom. Sci. Eng.*, vol. 20, no. 3, pp. 1983–1996, Jul. 2023.

[6] Y. Wu, X. Zhuo, L. Tang, W. Wu, and F. Qu, "Cooperative coverage path planning for AUVs in integrated underwater acoustic communication and detection networks," in *2023 IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2023, pp. 1–6.

[7] Y. Zhang, Q. Wang, Y. Shen, T. Wang, N. Dai, and B. He, "Multi-AUV cooperative search method based on dynamic optimal coverage," *Ocean Eng.*, vol. 288, no. 2, pp. 116 168–116 184, Nov. 2023.

[8] J. Zhang, G. Han, J. Sha, Y. Qian, and J. Liu, "AUV-assisted subsea exploration method in 6G enabled deep ocean based on a cooperative Pac-Men mechanism," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1649–1660, Feb. 2022.

[9] Y. Liu, J. Hu, and W. Dong, "Decentralized Coverage Path Planning with Reinforcement Learning and Dual Guidance," Oct. 2022, arXiv:2210.07514.

[10] R. A. McCarthy, S. T. Merrifield, J. Sarkar, and E. J. Terrill, "Reduced-order machine-learning model for transmission loss prediction in underwater acoustics," *IEEE J. Ocean. Eng.*, vol. 48, no. 4, pp. 1149–1173, Oct. 2023.

[11] L. M. Brekhovskikh, Y. P. Lysanov, and R. T. Beyer, *Fundamentals of ocean acoustics*. Springer, 1982.

[12] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, Jan. 2009.

[13] J. Xiao, Z. Chen, X. Sun, W. Zhan, X. Wang, and X. Chen, "Online multi-agent reinforcement learning for multiple access in wireless networks," *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3250–3254, Dec. 2023.

[14] J. M. Palacios-Gasós, E. Montijano, C. Sagüés, and S. Llorente, "Distributed coverage estimation and control for multirobot persistent tasks," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1444–1460, Dec. 2016.

[15] N. Basilico and F. Amigoni, "Exploration strategies based on multi-criteria decision making for searching environments in rescue operations," *Auton. Robot.*, vol. 31, pp. 401–417, Sep. 2011.